



Positivt og negativt samsvar

MEDISIN OG TALL

STIAN LYDERSEN

E-post: stian.lydersen@ntnu.no

Stian Lydersen (f. 1957) er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge), Institutt for psykisk helse, Fakultet for medisin og helsevitenskap, Norges teknisk-naturvitenskapelige universitet (NTNU).

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

To klinikere vil ikke alltid konkludere likt om hvorvidt en pasient har en diagnose. Cohens kappa er et mye brukt enkeltmål på samsvar. Men det kan være nyttig å skille mellom positivt samsvar og negativt samsvar – såkalt spesifikt samsvar. Dette er analogt til skillet mellom sensitivitet og spesifisitet for en diagnostisk test.

Hvis to klinikere vurderer samme pasient, vil de ikke alltid komme til samme konklusjon om pasientens diagnose. Hva er egnede mål på samsvar mellom observatørene?

Vi skal ta utgangspunkt i et eksempel hvor to psykologer, uavhengig av hverandre, satte diagnose basert på opptak av semistrukturerte intervju med barn og unge (1). Resultatene for opposisjonell atferdsforstyrrelse er vist i tabell 1. Klinikerne var enige i diagnosen i en andel $(25+2)/29 = 0,93$ av tilfellene. Justert for samsvar som forventes ved tilfeldigheter, er Cohens kappa lik 0,63, som bare så vidt regnes som godt samsvar (2). Men hvor relevant er det å oppsummere dette i bare ett tall? Nøyaktigheten til en diagnostisk test kvantifiseres jo med to størrelser: Sensitivitet, som er andel med positiv test blant de syke, og spesifisitet, som er andel med negativ test blant de friske (3). Tilsvarende kunne man definere positivt samsvar mellom observatører som andel enige blant pasienter som er syke, og negativt samsvar som andel enige blant pasienter som er friske. Men her kjenner vi ikke pasientens virkelige tilstand. Men de aktuelle tallene kan estimeres tilnærmet på følgende måte, der størrelsene a , b , c og d er som angitt i tabell 1. Vi antar at antall syke pasienter som blir korrekt klassifisert av begge observatørene, er tilnærmet lik d . Videre antar vi at halvparten av de pasientene som klassifiseres forskjellig av de to observatørene, er syke. Da vil totalt antall syke være $[d + (b + c)/2]$. Dermed får vi at positivt samsvar er $d/[d + (b + c)/2]$. Tilsvarende resonnering gir negativt samsvar: $a/[a + (b + c)/2]$.

Tabell 1

Samsvar mellom klinikere ved vurdering av om pasienten har diagnosen opposisjonell adferdsforstyrrelse (1)

Kliniker 2			
Kliniker 1	Nei	Ja	Totalt

Kliniker 2			
Kliniker 1	Nei	Ja	Totalt
Nei	25 (a)	2 (b)	27
Ja	0 (c)	2 (d)	2
Totalt	25	4	29

I eksempelet finner vi at positivt samsvar blir lik $2/[2 + (2+0)/2] = 0,67$, og negativt samsvar blir lik $25/[25 + (2 + 0)/2] = 0,96$. Vi ser at i dette tilfellet er man mer sikker på en negativ diagnose enn på en positiv diagnose.

Stor klinisk relevans

Et forslag til fortolkning av positivt samsvar er at dersom en kliniker klassifiserer pasienten som syk, så er dette sannsynligheten for at en annen kliniker, basert på samme informasjon, kommer til samme konklusjon (4). Tilsvarende kan man tolke negativt samsvar.

I motsetning til sensitivitet og spesifisitet vil positivt og negativt samsvar være avhengig av sykdommens prevalens blant dem som undersøkes. Samsvaret vil være størst i den største gruppen, typisk de friske, altså negativt samsvar. Jo lavere prevalensen er, jo lavere vil positivt samsvar bli (4).

Matematikken knyttet til negativt og positivt samsvar er behandlet av Feinstein (5) og Cicchetti (6). En kortfattet og velskrevet innføring finnes i artikkelen *Clinicians are right not to like Cohen's kappa* (4).

Spesifikt samsvar har to fordeler fremfor Cohens kappa (4, 7). For det første skiller det mellom positivt og negativt samsvar. Den kliniske konsekvensen av å bli klassifisert som syk er en annen enn konsekvensen av å bli klassifisert som frisk. For det andre: Ved lav prevalens og stor grad av enighet vil Cohens kappa alltid bli paradoksalt lav, noe som ikke er tilfelle i samme grad for spesifikke samsvar: Da vil negativt samsvar typisk bli høyt, mens positivt samsvar kan være moderat.

Til tross for at positivt og negativt samsvar har stor klinisk relevans, er målene ikke tatt mye i bruk ennå.

LITTERATUR:

1. Roelofs J, Muris P, Braet C et al. The Structured Clinical Interview for DSM-IV Childhood Diagnoses (Kid-SCID): first psychometric evaluation in a Dutch sample of clinically referred youths. *Child Psychiatry Hum Dev* 2015; 46: 367-75. [PubMed][CrossRef]
 2. Fagerland M, Lydersen S, Laake P. *Statistical analysis of contingency tables*. Boca Raton, FL: Chapman and Hall/CRC, 2017.
 3. Lydersen S. Hva er sannsynligheten for riktig resultat av en diagnostisk test? *Tidsskr Nor Legeforen* 2017; 137: 1437. [PubMed]
 4. de Vet HC, Mokkink LB, Terwee CB et al. Clinicians are right not to like Cohen's κ . *BMJ* 2013; 346 (apr12 1): f2125. [PubMed][CrossRef]
 5. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543-9. [PubMed][CrossRef]
 6. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551-8. [PubMed][CrossRef]
 7. de Vet HCW, Dikmans RE, Eekhout I. Specific agreement on dichotomous outcomes can be calculated for more than two raters. *J Clin Epidemiol* 2017; 83: 85-9. [PubMed][CrossRef]
-

Publisert: 19. mars 2018. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0963
Mottatt 6.11.2017, godkjent 5.1.2018.

